Statistical causes for the Epps effect in microstructure noise

Michael C. Münnix,* Rudi Schäfer, and Thomas Guhr

Department of Physics, University of Duisburg-Essen 47048 Dusiburg, Germany

We present two statistical causes for the distortion of correlations on high-frequency financial data. We demonstrate that the asynchrony of trades as well as the decimalization of stock prices has a large impact on the decline of the correlation coefficients towards smaller return intervals (Epps effect). These distortions depend on the properties of the time series and are of purely statistical origin. We are able to present parameter-free compensation methods, which we validate in a model setup. Furthermore, the compensation methods are applied to high-frequency empirical data from the NYSE's TAQ database. A major fraction of the Epps effect can be compensated. The contribution of the presented causes is particularly high for stocks that are traded at low prices.

I. INTRODUCTION

The decline of calculated correlations in financial data towards smaller return intervals was first discovered by Thomas Epps in 1979 [1]. This behavior was subsequently detected on different stock exchanges [2–4] and foreign exchange markets [5, 6]. The Epps effect has received considerable attention, from economists as well as from mathematicians and theoretical physicists.

Hayashi and Yoshida [7] introduced a cumulative estimator that only considers returns with overlapping time intervals. Hence, it deals with the asynchrony of time series as a cause for the Epps effect. Subsequently, Voev and Lunde [8] demonstrated that this estimator can be biased in the presence of noise and proposed a bias correction. Griffin and Oomen [9] extended the estimator of Hayashi and Yoshida by adjustments for lagged correlations. The work of Tóth and Kertész [10] also deals with the phenomenon of lagged correlations. They introduce a model that is based on the decomposition of cross-correlations. The recent study of Zhang [11], shows that usual previous-tick-estimators are biased. They consequently provide an optimal sampling frequency of returns to suppress the Epps effect. Barndorff-Nielson et. al. [12, 13] examine high frequency correlations and propose multivariate realized kernels to significantly improve the estimation of correlations. An extensive study of microscopic causes leading to the Epps effect has been performed by Renò [14].

Clearly, many mechanisms contribute to the Epps effect. We demonstrate that there are two major causes of purely statistical origin. Our aim is not to develop a complete description of the Epps effect. We rather want to identify statistical causes that can be compensated directly, without the requirement of adjusting parameters, model calibrations or an optimal sampling frequency. The two major causes we identify are the asynchrony of the time series and the impact of the decimalization by the tick-size.

This paper is organized as follows. In section II A and II B, we present compensation methods for the impact of asynchronous time series and the impact of the tick-size. This is followed by a combined compensation of both effects in section II C. The results are validated in a model setup in section III A. In section III B, we apply the compensation methods to empirical data from the NYSE's TAQ database to estimate the impact of the statistical causes on the Epps effect. We discuss our results in section IV.

II. THEORY OF COMPENSATING MICROSTRUCTURE NOISE DISTORTIONS

In the sequel, we give an overview over compensation methods that account for distortions of the Pearson correlation coefficient due to statistical effects. In particular, the asynchrony of trading times and the impact of the tick-size are considered.

A. Asynchrony of trading times

We begin with demonstrating how the asynchrony of time series contributes to the Epps effect. By asynchrony we refer to time series that feature an arbitrary lag for a given point in time but the average lag is zero. The asynchrony is simply due to the non-synchronous pricing of stocks. A detailed derivation and study of our finding is performed in [15]

Tóth and Kertész [10] stated that the impact of the asynchrony is weak, compared to the impact of a static lag, for which they developed a model. In the following, we will demonstrate that the asynchrony can be a major cause for the Epps effect.

The central assumption of this approach is the existence of underlying non-lagged time series of prices. The assumption of a finer (see, e.g., [10]) or continuous (see, e.g., ([7, 12, 14]) timescale is frequently used in the estimation of correlations. This ansatz is also intuitive, as most stocks are traded at several stock exchanges and off-exchange (OTC) simultaneously. The basic idea is the following: Due to the asynchrony, each term of the Pearson correlation coefficient can be divided into a part which contributes to the correlation and a part which is
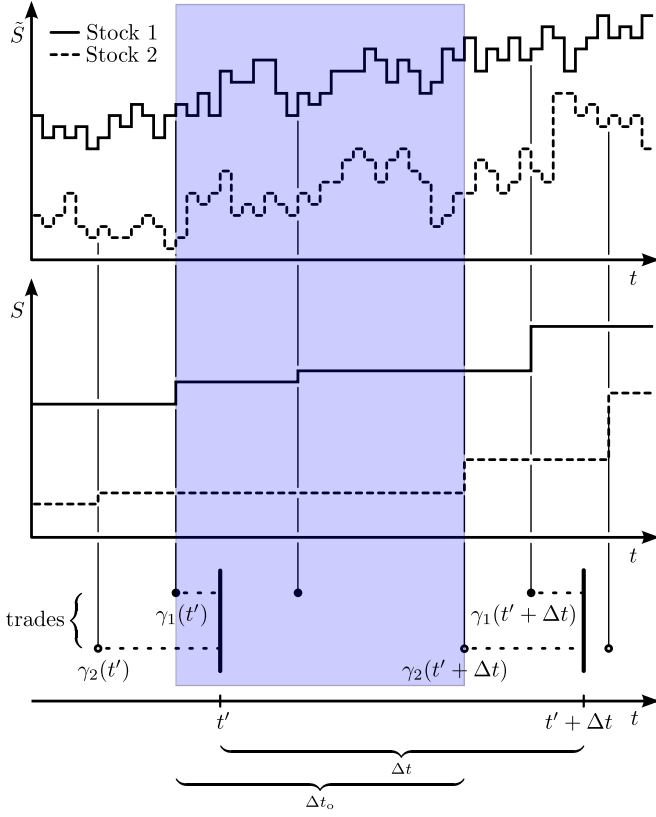
FIG. 1. Illustration of the model for asynchronous trading times of two stocks. Shown on the top are the prices $\tilde{S}$ for the hypothetical underlying timescale. The "sampling" of theses prices to macroscopic prices S with randomly distributed points of trades is illustrated in the middle. The points of trades are indicated by the vertical lines. The thick bars at the bottom illustrate the return interval between t' and $t'+\Delta t$. The points of last trades on these times are denoted with γ. The shaded area indicates the overlap Δt_{o}.

uncorrelated and therefore lowers the correlation coefficient.

The situation is sketched in Fig. 1. Here, $\gamma_i(t)$ refers to the point of last trade, for the i-th stock at time t. The waiting times, that is, the periods between two consecutive trades, are randomly (usually exponentially) distributed. Hence, when calculating the return of the interval from t' to $t' + \Delta t$, we actually obtain the return on an effective return interval which is between the points of last trade referring to the right and the left side of the return interval, that are $[\gamma_1(t'); \gamma_1(t' + \Delta t]$ and $[\gamma_2(t'); \gamma_2(t' + \Delta t]$. These intervals can be smaller or larger than the initially chosen return interval. When considering the returns of two stocks within the same interval, one obtains two effective return intervals that are in most cases not equal in length, start-point and end-point. These intervals usually share an overlap Δt_{o}, although this is not necessarily true for stocks that are

traded on low quantities. This overlap is given by

$$
\begin{aligned}
\Delta t_{\mathrm{o}}(t') = {} & \min\left(\gamma_1(t' + \Delta t), \gamma_2(t' + \Delta t)\right) \\
& - \max\left(\gamma_1(t'), \gamma_2(t')\right) \ .
\end{aligned}
\tag{1}
$$

The fractional overlap $\Delta t_{\mathrm{o}}/\Delta t$ declines with lower return intervals as shown in Fig. 2. This scaling behavior already looks similar to the Epps effect on correlation coefficients. As we will demonstrate, the fractional overlap is strongly connected to the Epps effect.

In the sequel, we consider the Epps effect on relative price changes or arithmetic financial returns r which are defined as the relative price change during a return interval Δt. It reads

$$
r(t) = \frac{S(t + \Delta t) - S(t)}{S(t)} \ ,
\tag{2}
$$

where $S(t)$ refers to the price of a security at time t.

Regarding the hypothetical underlying time series, the information within the overlap Δt_{o} is synchronous. This part gives the *true* correlation of the time series. In contrast, the parts left and right from the overlap are asynchronous. Under the assumption of randomly distributed trading times, these parts are on average uncorrelated. Hence, the returns outside the overlap distort the correlation coefficient.

It follows that the contribution of these two single returns to the Pearson correlation coefficient is the Pearson correlation coefficient of the underlying time series multiplied by the fractional overlap as shown in [15]. This can easily be outlined, when considering the normalized returns of the underlying time series

$$
\tilde{g}(t) = \frac{\tilde{r}(t) - \langle \tilde{r} \rangle}{\sigma_{\tilde{r}}} \ .
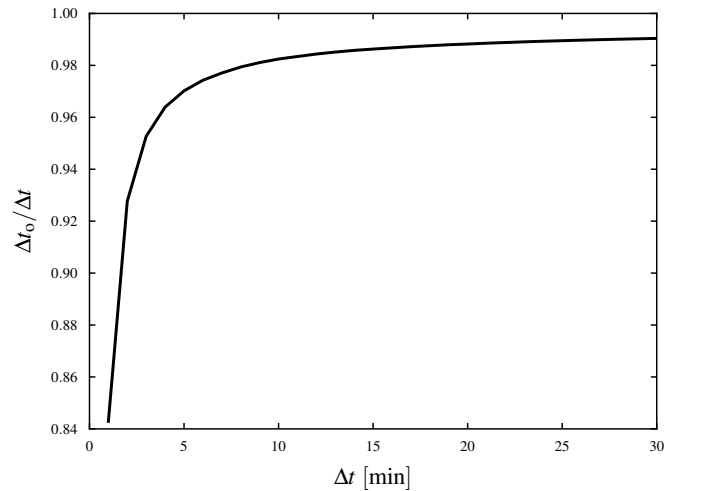\tag{3}
$$

FIG. 2. Average fractional overlap for the 5 highest correlated stock pairs of each industry branch in the S&P 500 index versus the return interval Δt.

Here, $\langle\cdots\rangle$ denotes the average over T and σ denotes to the standard deviation of the time series with length T. The tilde indicates the underlying time series. When calculating the Pearson correlation coefficient of two time series, the overlap is a function of the time step: $\Delta t_o = \Delta t_o(t)$. We denote the interval of the overlap $\Delta t_o(t)$ for each time step as $\mathcal{J}(t)$. The time steps on the underlying timescale are denoted with $\tilde{t}$. We can re-arrange the terms of the correlation coefficient in terms that originate from within the overlap-interval and thus are synchronous, and terms that are asynchronous,

$$
\mathrm{corr}(r_1, r_2) \propto \frac{1}{T} \sum_{t=0}^{T} \left(\left(\underbrace{\sum_{\tilde{t} \notin \mathcal{J}(t)} \tilde{g}_1(\tilde{t})}_{\text{async.}} + \underbrace{\sum_{\tilde{t} \in \mathcal{J}(t)} \tilde{g}_1(\tilde{t})}_{\text{sync.}} \right) \times \left(\underbrace{\sum_{\tilde{t} \notin \mathcal{J}(t)} \tilde{g}_2(\tilde{t})}_{\text{async.}} + \underbrace{\sum_{\tilde{t} \in \mathcal{J}(t)} \tilde{g}_2(\tilde{t})}_{\text{sync.}} \right) \right) . \quad (4)
$$

This leads to

$$
\mathrm{corr}(r_1, r_2) = \frac{1}{T} \sum_{t=0}^{T} \mathrm{corr}_t(\tilde{g}_1, \tilde{g}_2) \frac{\Delta t_o(t)}{\Delta t} , \quad (5)
$$

where, $\mathrm{corr}_t(\tilde{g}_1, \tilde{g}_2)$ is the Pearson correlation coefficient of the underlying time series for the interval $[t, t + \Delta t]$. It gives the *true* correlation. Each term of the correlation coefficient is multiplied by the fractional overlap $\Delta t / \Delta t_o(t)$, because only the information inside the overlap contributes to the correlation coefficient.

As we are able to quantify the impact of the fractional overlap on the correlation coefficient, we can easily compensate this distortion by

$$
\widehat{\mathrm{corr}}_{\text{async}}(r_1, r_2) = \left\langle g_1(t_j) g_2(t_j) \frac{\Delta t}{\Delta t_o(t_j)} \right\rangle , \quad (6)
$$

where g refers to the normalized return of the corresponding (non-hypothetical) time series. Furthermore, only returns should be considered that actually share an overlap, analogously to the estimator of Hayashi and Yoshida [7].

Initially, we made the assumption of an underlying time series of prices, which is correlated and which exists on a smaller time scale. Equation (6) does no longer depend on the time scale of the hypothetical underlying time series. Neither does it depend on the actual prices on the underlying time series. Hence, the only necessary assumption is that there exists underlying *information* which is correlated on a finer time scale. This is an important finding, since Martens and Poon [16] indicated that the synchronization of returns from international stock exchanges is a non-trivial problem.
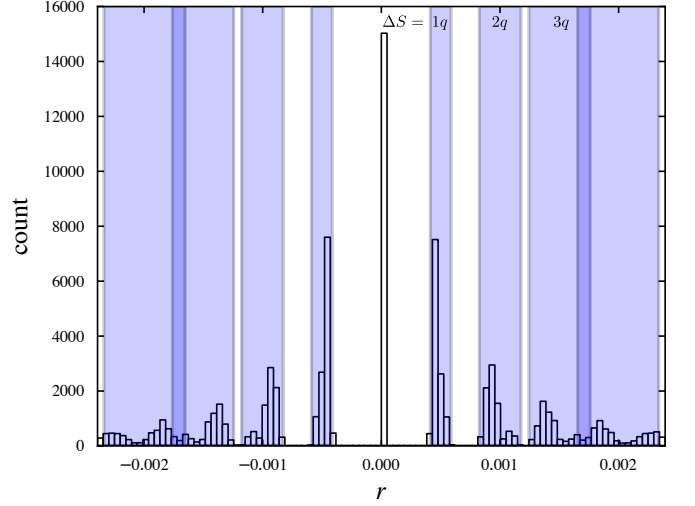

FIG. 3. Detail of the distribution of 5-min returns of the AES Corp. share in 2007. The shaded areas refer to returns that originate from the same absolute price change ΔS. The price changes ΔS are denoted as multiplies of the tick-size q.

B. Tick-Size

We now turn to the second statistical cause. We estimate the tick-size's impact on the Epps effect. A comprehensive derivation and discussion is performed in [17].

The lowest possible price change, the *tick-size*, of most securities has been constantly reduced, resulting in tick-sizes of 1/100-th of the respective currency on most stock exchanges. This process is often referred to as decimalization. It was, e.g., motivated by aiming at an enhanced market efficiency. In theory, small tick-sizes allow for a faster clearing of market arbitrage. However, the question whether a smaller tick-size generally improves the market quality is controversially discussed [18, 19]. Among others, Harris [20] discussed that a larger tick-size can ensure liquidity, but on the other hand, it can lead to erroneous data in financial indices [21]. In this context, Angel [22] observed that the prices of a stock are commonly in a typical range, which is optimal to provide liquidity. Companies perform stock splits to control the absolute price of their share. A recent study by Onnela et. al. [23] indicates that in some cases only a fraction of the theoretically possible prices are used. Hence, prices cluster at certain multiples of the tick-size resulting in an effective tick-size.

At first glance, it could appear that the transition from absolute price changes ΔS to returns r removes this discretization from the distribution, since the returns are almost continuously distributed. A detailed look at the center of a return distribution (see Fig. 3) reveals that the discretization effects are still visible. Despite its poor graphical visibility, the discretization affects returns on all intervals. Especially, we expect an impact on the correlation coefficient if the discretization is high, that is,

when stocks are traded at low prices.

Due to the imposing of discrete prices, information is lost. The average relative price change becomes smaller when considering smaller return intervals. The tick-size remains the same of course. The information loss grows for smaller return intervals. Thus, the discretization should also contribute to the Epps effect.

The basic assumption of our model is that we can statistically describe the discreteness in market prices by a discretization of a hypothetical underlying price. Of course, the market prices do not actually result from a discretization process. However, there is a large variety of trading strategies simultaneously acting on the market. These strategies also act on a large spectrum of different investment horizons. There are even traders that try to exploit the finite tick-site in their trading strategies. As the price formation results from the interaction of this diversity of trading strategies, the price fluctuations on the level of the tick-size can be viewed as purely statistical. Hence, a natural approach is the assumption of on average uniformly distributed discretization errors.

Using the arithmetic return defined in equation (2), we introduce the discretization error ϑ as

$$\bar{r}(t) = \frac{\Delta \bar{S}(t)}{\bar{S}(t)} \tag{7}$$

$$r(t) = \frac{\Delta \bar{S}(t) + \vartheta}{\bar{S}(t)} \tag{8}$$

with

$$\Delta \bar{S}(t) = \bar{S}(t) - \bar{S}(t + \Delta t) , \tag{9}$$

where $\bar{S}(t)$ denotes to the discretized stock price and $\bar{r}(t)$ denotes the return, which is based on discretized stock prices. We emphasize that we do not account for the discretization of the prices $S(t)$, but consider only the discretization of the price changes $\Delta S(t)$. We demonstrate in section III A that this only induces a negligible error. As ϑ in equation (8) is actually the difference of two uniformly distributed discretization errors, it follows a triangular distribution.

The calculation of the Pearson correlation coefficient including the discretization errors as introduced in (8) leads to

$$\widehat{\text{corr}}_{\text{tick}}(r_1, r_2) = \frac{\text{cov}(r_1, r_2)}{\sigma_{r_1} \sigma_{r_2}} \tag{10}$$

$$= \Bigg[\text{cov}\left(\bar{r}_1, \bar{r}_2 \right)$$

$$\text{cov}\left(\frac{\Delta \bar{S}_1}{S_1}, \frac{\vartheta_2}{S_2} \right) + \text{cov}\left(\frac{\Delta \bar{S}_2}{S_2}, \frac{\vartheta_1}{S_1} \right)$$

$$+ \text{cov}\left(\frac{\vartheta_1}{S_1}, \frac{\vartheta_2}{S_2} \right) \Bigg] \Bigg/ \left(\hat{\sigma}_{r_1} \hat{\sigma}_{r_2} \right) \tag{11}$$

with

$$\hat{\sigma}_{r_i} = \sqrt{ \text{var}\left(\bar{r}_i \right) + \text{var}\left(\frac{\vartheta_i}{S_i} \right) + 2\text{cov}\left(\frac{\Delta \bar{S}_i}{S_i}, \frac{\vartheta_i}{S_i} \right) } . \tag{12}$$

Here, $\bar{r}$ refers to the return with respect to discretized prices. Only the terms $\text{cov}\left(\bar{r}_1, \bar{r}_2 \right)$, $\text{var}\left(\bar{r}_1 \right)$ and $\text{var}\left(\bar{r}_2 \right)$ can be calculated with the discretized prices. All other terms are unknown and describe the information loss due to the discretization. We can estimate these terms and thereby compensate for the information loss by interpolating the price change distribution. Technically, this is achieved by expanding the variance and covariance terms in equation (11) and (12) and estimating the discretization error for all price changes individually. Estimation techniques for the individual discretization errors are comprehensively discussed in [17]. This study indicates that only certain terms of equation (11) have a noticeable impact on the compensation. If calculation speed is an issue, one can approximate

$$\widehat{\text{corr}}_{\text{tick}}(r_1, r_2) \approx \frac{\text{cov}\left(\bar{r}_1, \bar{r}_2 \right)}{\hat{\sigma}_{r_1} \hat{\sigma}_{r_2}} . \tag{13}$$

The main contribution to the distortion of correlation coefficients in small return intervals is the overestimation of σ. Fig. 7 in section III A shows this overestimated σ and the tick-size-corrected $\hat{\sigma}$ versus the return interval Δt. This is consistent with the findings of Hansen and Lunde [24]. They demonstrate that the realized variance is overestimated on small return intervals due to microstructure noise. The empirical evidence in section III B indicates that the tick-size have profound impact on this noise.

Due to the convex shape of the price change distribution, the discretization errors are not distributed symmetrically. This effect grows with the impact of the discretization, i.e., smaller return intervals. Thus, the estimation of variances on the discretized values is biased. This gives the largest contribution to the distortion of correlation coefficients due to discretized data. We can correct this behavior with the presented compensation.

C.   Combined compensation

Having presented compensation methods for distortions of the correlation coefficient due to asynchronous time series and due to the tick-size, we now combine both findings. The compensation of asynchrony acts on each term of the Pearson correlation coefficient for every point in time. The tick-size compensation, in contrast, acts on the Pearson correlation coefficient as a whole in terms of the time series, but it acts on every occurring price change individually. Both effects superimpose, as illustrated in Fig. 4. The horizontal axis shows the product of normalized 1-min returns for each point in 2007 (overnight returns are excluded). The vertical axis shows the corresponding fractional overlap of each return pair. The discretization effects are visible in the center, superimposed with the asynchronous characteristics. Similar to the findings of Szpiro [25] for single stocks, the tick-size induces a nanostructure on the terms of the Person correlation coefficient for two stocks.
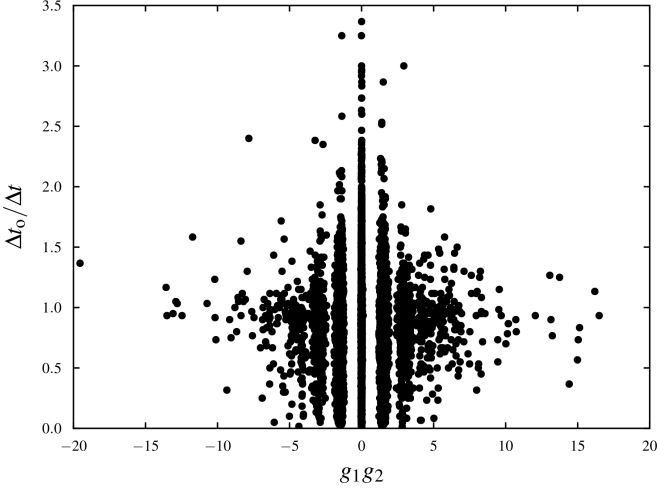
FIG. 4. Product of normalized return pairs versus fractional their fractional overlap for 1-min returns of the shares of Novell Inc. and Unisys Corp.in 2007. The average fractional overlap is 0.76.

FIG. 5. Model results of compensation methods.

The simultaneous compensation of both effects can be achieved by combining both presented compensations. It reads

$$\widehat{\mathrm{corr}}(r_1, r_2) = \left\langle \bar{r}_1 \bar{r}_2 \frac{\Delta t}{\Delta t_\mathrm{o}} \right\rangle \tag{14}$$

$$+ \left(\mathrm{cov}\left(\frac{\Delta \bar{S}_1}{S_1}, \frac{\vartheta_2}{S_2} \right) + \mathrm{cov}\left(\frac{\Delta \bar{S}_2}{S_2}, \frac{\vartheta_1}{S_1} \right) \right.$$

$$+ \mathrm{cov}\left(\frac{\vartheta_1}{S_1}, \frac{\vartheta_2}{S_2} \right) - \langle \bar{r}_1 \rangle \langle \bar{r}_2 \rangle \Bigg)$$

$$\times \left\langle \frac{\Delta t}{\Delta t_\mathrm{o}} \right\rangle \Big/ (\hat{\sigma}_{r_1} \hat{\sigma}_{r_2}) \ . \tag{15}$$

Analogously to the previous section, this expression can be approximated by

$$\widehat{\mathrm{corr}}(r_1, r_2) \approx \frac{\left\langle \bar{r}_1 \bar{r}_2 \frac{\Delta t}{\Delta t_\mathrm{o}} \right\rangle}{\hat{\sigma}_{r_1} \hat{\sigma}_{r_2}} \ . \tag{16}$$

By multiplying the covariance terms of discretized returns $\bar{r}$ with the inverse fractional $\Delta t/\Delta t_\mathrm{o}$ overlap and by correcting the overestimation of the standard deviations σ, the largest fraction of the correlation coefficient's distortion can be compensated.

III. RESULTS

Before applying the method to empirical data, we study it in a model setup. Subsequently, we apply the compensation methods to empirical data from the NYSE's TAQ database to estimate the impact of the presented causes on the distortion of correlation coefficients.
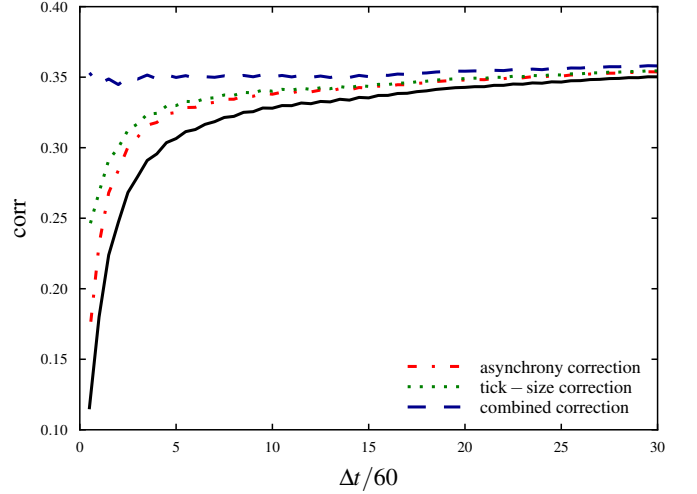
A. Model results

We start by generating an underlying correlated time series using a GARCH(1,1) model, as introduced in [26]

$$r_i(t) = \sigma_i(t) \left(\sqrt{c}\,\eta(t) + \sqrt{1-c}\,\varepsilon_i(t) \right) \ . \tag{17}$$

Here $r_i(t)$ stands for the return of the i-th stock at time t and c is the correlation coefficient. The random variables $\eta(t)$ and $\varepsilon_i(t)$ are taken from standard normal distributions. $\eta(t)$ is identical for all stocks; It induces the correlation. The ε_i are individual for each stock. $\sigma_i(t)$ is the non-constant variance, given by a GARCH(1,1) process

$$\sigma_i^2(t) = \alpha_0 + \alpha_1 r_i^2(t-1) + \beta_1 \sigma_i^2(t-1) \ . \tag{18}$$

The initial parameters of the GARCH(1,1) process are chosen as $\alpha_0 = 2.4 \times 10^{-4}$, $\alpha_1 = 0.15$ and $\beta_1 = 0.84$.

Two return time series r_1 and r_2 are generated representing two correlated stocks. The total lengths of these time series is chosen as 7.2×10^6, corresponding to a return interval Δt of one second during one trading year. From these returns, we generate two underlying price time series $\tilde{S}_1$ and $\tilde{S}_2$. We set the starting prices to $t = 0$ to 1000. c is chosen as 0.4.

To model the asynchronous trade processes, these prices are sampled independently using exponentially distributed waiting times with average values typical for the stock market. We choose the average waiting times as 15 and 25 data points (equivalent to seconds in this setup). In the next step, we round the prices to integer values. An integer price of, for example, 1000 then corresponds to a price of 10 and a tick-size of 0.01.

Finally, we construct the time series of returns from these prices using return intervals from 60 data points (corresponding to 1 minute) to 1800 data points (corresponding to 30 minutes). The thus obtained time series features both, asynchrony and discretization.

The results of the applied compensation methods are shown in Fig 5. We are able to compensate the statistical distortion of correlation coefficients almost completely. The remaining decline of the corrected correlation coefficient on very small return intervals is due to the approximations presented in sections II B and II C (only price change discretization is considered) as well as the negligence of the correlation between price changes and prices and the discretization of prices. The impact of the overestimation of the standard deviation σ is shown in Fig. 7. This illustrates that the tick-size can have a large impact on the overestimation of σ. Moreover, we are able to compensate for this behavior down to approximately $\Delta t = 180$ time steps (corresponding to 3 minutes in our model).

B. Empirical evidence

As already mentioned, many mechanisms contribute to the Epps effect. Our present aim is to quantify the part, which is caused by the statistical properties of the time series.

It is difficult to isolate the Epps effect on single stock pairs, as it is superimposed with other effects leading to other characteristics of the correlation coefficient than expected for the Epps effect.

Because of that, we classify two ensembles of stock pairs. After compensating the asynchrony effect for each pair, we build the average for the ensemble by normalizing the correlation coefficients individually by their saturation value at a return interval of 30 minutes. We also plot the error bars of the compensation representing the double standard deviation 2σ of the correction. By this method, we can show the scope of the asynchrony model and identify regions, in which other effects dominate. All data is extracted from the NYSE's TAQ database for the year 2007.

The first ensemble consists of 5 stock pairs of each industry sector of the S&P 500 index (50 stocks in total), whose daily returns provide the strongest correlation during the year 2007. We applied the asynchrony compensation to this ensemble. The results shown in Fig 6(a) indicate that the asynchrony has a pronounced impact on the Epps effect. It appears that asynchrony effects are the dominating cause for the Epps effect on return intervals down to approximately 10 minutes, where the remaining Epps effect is on average less than 3% of the correlation coefficient's saturation value at large return intervals. Of course, within the statistical ensemble stock pairs can be found which either do not show an Epps effect or which are so infrequently traded that the assumption of an underlying timeline might be unreasonable. Even though the assumption of an underlying time series is a common and intuitive approach, it may not be valid for stocks traded on very low frequencies.
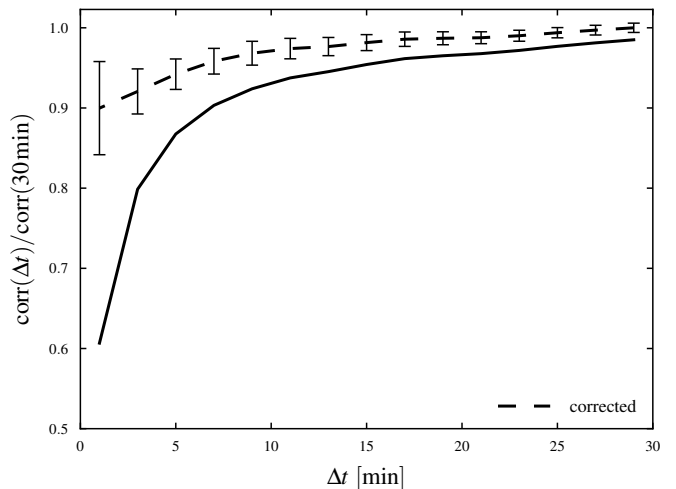
The second ensemble serves as a test scenario for the tick-size compensation. We expect the tick-size to only
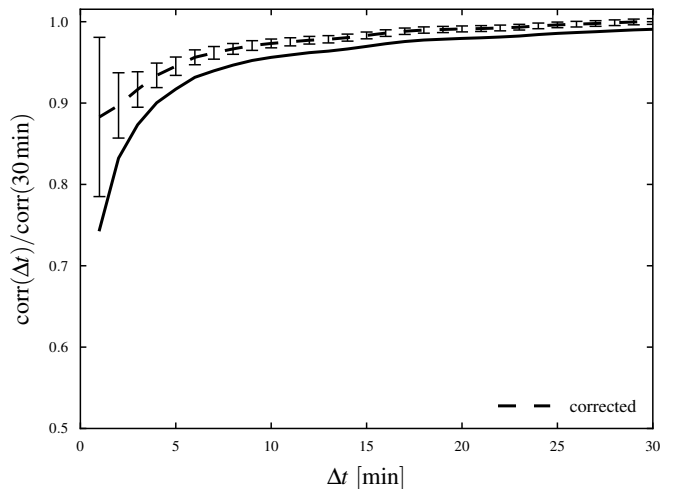
(a)Asynchrony compensation
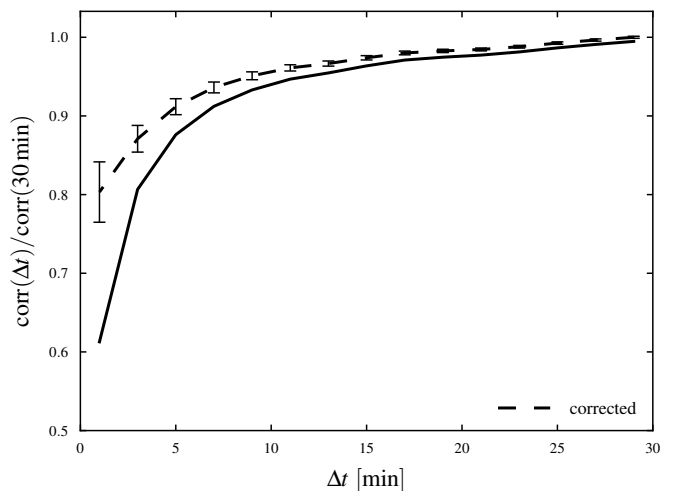
(b)Tick-size compensation

(c)Combined compensation

FIG. 6. Empirical results of applied compensation methods. a) represents the average over the 50 highest correlated stock pairs in the S&P 500 index in 2007 (Top 5 from each industry branch). b) and c) represent the average over the 25 highest correlated stock pairs that were traded between \$10.01 and \$20.00 in 2007. The correlation coefficients have been individually normalized to the corrected value at $\Delta t = 30$ min
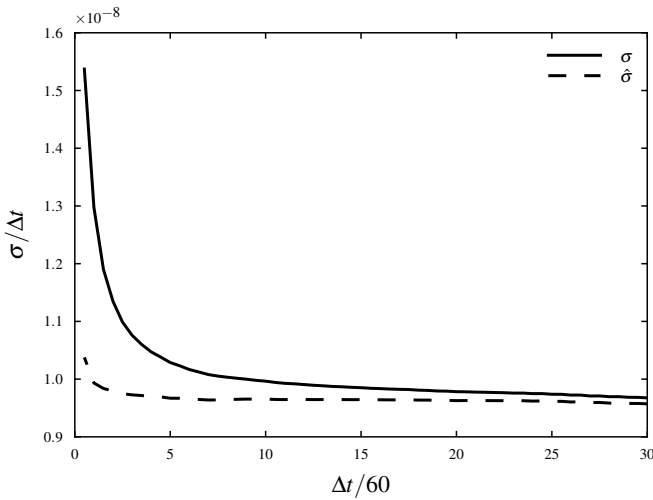
FIG. 7. Overestimated standard deviation σ and tick-size-corrected standard deviation $\hat{\sigma}$ versus the return interval Δt within the model.

have a large impact on the correlation coefficient, if the discretization is high, i.e., if stocks are traded at low prices. Thus, we construct the second ensemble out of the 25 most strongly correlated stocks in the S&P 500 that are traded in the range of \$10.01 to \$20.00. The price change distributions are segment-wise interpolated with heavy tailed distributions as, i.e., suggested by [27]. The results in Fig. 6(b) indicate that for stocks that are traded at low prices, the tick-size can have a sizable impact on the Epps effect.

Eventually, we apply a combined compensation to the second ensemble. Results are shown in Fig. 6(c). The empirical evidence indicates that statistical effects can have a very profound impact on the Epps effect.

IV. CONCUSIONS

We demonstrated that statistical causes can have a large impact on the Epps effect, especially for stocks that are traded at low prices. The asynchrony of time series as well as the tick-size have a major impact on the Epps effect. We developed two simple methods to compensate for these causes.

However, this is not a full description of the Epps effect as there are certainly many phenomena contributing to it. In certain scenarios, other statistical properties of the time series or other causes for the Epps effect might dominate. The size of the error bars in Fig. 6 indicates that the asynchrony compensation does not give reliable results for return intervals below 3 minutes. Especially for very small return intervals, a lag between the time series of two stocks might be the dominating cause, as suggested by Tóth and Kertész [10].

For stocks that are infrequently traded at very low prices (often referred to as *penny-stocks*) the assumption of uniformly distributed discretization errors needs to be carefully reflected. It is possible that certain trading strategies dominate for those stocks leading to an asymmetrical distribution of discretization errors.

Nonetheless, the presented compensations significantly improve the estimation of financial correlations. These methods do not require parameter adjustments or model calibrations. Our empirical study indicates that the identified causes can contribute up to 75% of the Epss effect for stocks that are traded at low prices.

[1] T. W. Epps, Comovements in stock prices in the very short run, Journal of the American Statistical Association 74 (366) (1979) 291–298.
URL http://www.jstor.org/stable/2286325

[2] G. Bonanno, F. Lillo, R. N. Mantegna, High-frequency cross-correlation in a set of stocks, Quantitative Finance 1 (1) (2001) 96–104.

[3] J. Kwapien, S. Drozdz, J. Speth, Time scales involved in emergent market coherence, Physica A 337 (1-2) (2004) 231 – 242. doi:10.1016/j.physa.2004.01.050.
URL http://www.sciencedirect.com/science/article/B6TVG-4BRJY2K-1/2/d141755cb92c5d1b33cead1bb57b0bed

[4] A. A. Zebedee, M. Kasch-Haroutounian, A closer look at co-movements among stock returns, Journal of Economics and Business 61 (4) (2009) 279 – 294. doi:10.1016/j.jeconbus.2008.11.001.
URL http://www.sciencedirect.com/science/article/B6V7T-4TX794N-1/2/26edc1f881adc4aad35511cb584cb8d2

[5] M. Lundin, M. Dacorogna, U. A. Müller, Financial Markets Tick By Tick, Wiley, 1998, Ch. Correlation of High Frequency Financial Time Series.

[6] J. Muthuswamy, S. Sarkar, A. Low, E. Terry, Time variation in the correlation structure of exchange rates: high-frequency analyses, Journal of Futures Markets 21 (2)

(2001) 127–144.

[7] T. Hayashi, N. Yoshida, On covariance estimation of non-synchronously observed diffusion processes, Bernoulli 11 (2) (2005) 359–379.

[8] V. Voev, A. Lunde, Integrated Covariance Estimation using High-frequency Data in the Presence of Noise, Journal of Financial Econometrics 5 (1) (2007) 68–104. `arXiv:http://jfec.oxfordjournals.org/cgi/reprint/5/1/68.pdf`, `doi:10.1093/jjfinec/nbl011`. URL `http://jfec.oxfordjournals.org/cgi/content/abstract/5/1/68`

[9] J. E. Griffin, R. C. Oomen, Covariance measurement in the presence of non-synchronous trading and market microstructure noise, Journal of Econometrics In Press, Corrected Proof.

[10] B. Tóth, J. Kertész, The epps effect revisited, Quantitative Finance 9 (7) (2009) 793–802. `doi:10.1080/14697680802595668`.

[11] L. Zhang, Estimating covariation: Epps effect, microstructure noise, Journal of Econometrics In Press. `doi:10.1016/j.jeconom.2010.03.012`.

[12] O. E. Barndorff-Nielsen, N. Shephard, Econometric analysis of realised covariation: High frequency covariance, regression and correlation in financial economics, Economics Papers 2002-W13, Economics Group, Nuffield College, University of Oxford (2001). URL `http://ideas.repec.org/p/nuf/econwp/0213.html`

[13] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, N. Shephard, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, Economics Papers 2008-W10, Economics Group, Nuffield College, University of Oxford (2008). URL `http://ideas.repec.org/p/nuf/econwp/0810.html`

[14] R. Renò, A closer look at the epps-effect, International Journal of Theoretical and Applied Finance 6 (1) (2003) 87–102.

[15] M. C. Münnix, R. Schäfer, T. Guhr, Compensating asynchrony effects in the calculation of financial correlations, Physica A 389 (4) (2010) 767–779. `doi:10.1016/j.physa.2009.10.033`.

[16] M. Martens, S.-H. Poon, Returns synchronization and daily correlation dynamics between international stock markets, Journal of Banking & Finance 25 (10) (2001) 1805 – 1827. `doi:10.1016/S0378-4266(00)00159-X`. URL `http://www.sciencedirect.com/science/article/B6VCY-43XFHT0-1/2/d81bc0d098bd245deef0e9d31f352682`

[17] M. C. Münnix, R. Schäfer, T. Guhr, Impact of the tick-size on financial returns and correlations, Physica A 389 (21) (2010) 4828–4843, http://dx.doi.org/10.1016/j.physa.2010.06.037. `doi:10.1016/j.physa.2010.06.037`.

[18] D. Bourghelle, F. Declerck, Why markets should not necessarily reduce the tick size, Journal of Banking & Finance 28 (2) (2004) 373 – 398. `doi:10.1016/S0378-4266(03)00136-5`.

[19] R. D. Huang, H. R. Stoll, Tick size, bid-ask spreads, and market structure, Journal of Financial and Quantitative Analysis 36 (04) (2001) 503–522. `doi:10.2307/2676222`.

[20] L. Harris, Stock price clustering and discreteness, Review of Financial Studies 4 (3) (1991) 389–415. `doi:10.1093/rfs/4.3.389`.

[21] S. Kozicki, B. Hoffman, Rounding error: A distorting influence on index data, Journal of Money, Credit and Banking 36 (3) (2004) 319–38. URL `http://ideas.repec.org/a/mcb/jmoncb/v36y2004i3p319-38.html`

[22] J. J. Angel, Tick size, share prices, and stock splits, Journal of Finance 52 (2) (1997) 655–681. URL `http://ideas.repec.org/a/bla/jfinan/v52y1997i2p655-81.html`

[23] J.-P. Onnela, J. Töyli, K. Kaski, Tick size and stock returns, Physica A: Statistical Mechanics and its Applications 388 (4) (2009) 441 – 454. `doi:10.1016/j.physa.2008.10.014`. URL `http://www.sciencedirect.com/science/article/B6TVG-4TPPF19-1/2/282ce63e7c76e54d433dc58fb164079f`

[24] P. R. Hansen, A. Lunde, Realized variance and market microstructure noise, Journal of Business & Economic Statistics 24 (2006) 127–161. URL `http://ideas.repec.org/a/bes/jnlbes/v24y2006p127-161.html`

[25] G. G. Szpiro, Tick size, the compass rose and market nanostructure, Journal of Banking & Finance 22 (12) (1998) 1559 – 1569. `doi:DOI:10.1016/S0378-4266(98)00073-9`. URL `http://www.sciencedirect.com/science/article/B6VCY-41XMJV1-4/2/3bfb8017f7acde3b7cebc5d474314946`

[26] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (3) (1986) 307 – 327. `doi:10.1016/0304-4076(86)90063-1`. URL `http://www.sciencedirect.com/science/article/B6VC0-46VV78N-4/2/ce371daca28a0d38824736988b1d0ef1`

[27] R. Cont, P. Tankov, Financial Modelling with Jump Processes, second edition Edition, CRC Press, Taylor and Francis, 2010, Ch. 7.